3П

УДК 004.934

В.В. Пилипенко, О.Н. Ладошко
Международный научно-учебный центр информационных технологий и систем,
г. Киев, Украина
valery_pylypenko@mail.ru
Национальный политехнический университет Украины «Киевский политехнический институт», г. Киев, Украина
ladoshko@gmail.com

Аннотация и учет речевых сбоев в задаче автоматического распознавания спонтанной украинской речи

В статье рассматривается проблема влияния речевых сбоев на процесс автоматического распознавания спонтанной украинской речи на примере автоматизированного стенографа для получения текста стенограммы из звукового файла. Записанная фонограмма обрабатывается системой распознавания слитной речи многих дикторов из больших словарей (больше 10 тыс. слов). Рассматриваются и систематизируются основные типы речевых сбоев в спонтанной украинской речи. На основе учета речевых сбоев и коррекции стенограммы производится очистка данных, что позволяет улучшать показатели надёжности распознавания речи.

Введение

Автоматическое распознавание спонтанной речи усложняется за счет различного рода нарушений плавного развёртывания речевого потока. Примером таких нарушений могут быть самоисправления, колебания, вокализованные паузы, заполненные элементами «е», «а», «м-м». Такие явления объединяют в класс речевых сбоев (дисфлуенций) [1], мешающих процессу автоматического распознавания спонтанной речи. Одним из классов прикладных задач автоматического распознавания спонтанной речи является стенографирование. Системы стенографирования предназначены для автоматизированной обработки материалов заседаний различных форумов. Использование автоматического преобразования речевого сигнала в текст позволяет значительно упростить работу стенографиста. В корпусах украинской речи ранее не производилось выявление и изучение влияния речевых сбоев на надежность автоматического распознавания речи. Это можно объяснить тем, что ручное аннотирование речевых сбоев аудиозаписи и их отображение в транскрипте – задача, требующая затрат времени и квалифицированной экспертной работы. Тем не менее, в корпусах ограниченного объёма решение такой задачи является посильным делом. Ручная разметка речевых сбоев в транскрипте стенограммы повышает информативное содержание речевого корпуса, а также расширяет круг вопросов, решаемых с помощью корпуса.

Целью данной работы является аннотирование и учет основных речевых сбоев спонтанной украинской речи для улучшения показателей надёжности системы автоматического распознавания.

1 Речевые сбои спонтанной речи

Характерной особенностью спонтанной речи является наличие неинформативных элементов. Чаще всего неинформативные элементы, так называемые речевые сбои [2], – это звуки, производимые самим диктором либо же звуки, вызванные

нарушениями внешних условий коммуникации (вмешательство собеседника). Эти неинформативные элементы следует рассматривать как помеху, которая искажает исходный языковой материал.

Можно отметить следующие основные причины возникновения речевых сбоев дикторов:

– несоблюдение орфоэпических норм. Эти нормы включают правильное произношение (артикуляция) отдельных звуков, звукосочетаний, нормативное ударение слов и их форм;

– ускоренный темп речи. Из-за ограничения времени выступлений речь многих депутатов имеет быстрый темп;

– не всегда равномерный и устойчивый поток воздуха, который выдыхается диктором в процессе монолога, так называемое «фонационное дыхание»;

– не всегда интонационно логично выделены важные моменты речи. Логически-интонационная речь должна выделяться паузами. Такие паузы отделяют одну фразу от другой;

– уменьшение громкости произношения окончаний слов. Это явление вызывает появление элементов речи, не учтённых в лексиконе системы автоматического распознавания;

– отступление от подготовленного доклада вызывает речевые сбои, связанные с внутренними проблемами планирования и развёртывания дискурса.

В результате из-за речевых сбоев фраза диктора становится «зашумленной» избыточной информацией [3], мешающей не только процессу проистекания и понимания речи, но и процессу их автоматического распознавания.

В работе [4] отмечают, что число речевых сбоев в монологах и диалогах неформальной разговорной речи может достигать 15 – 20% всех произнесенных слов. Поэтому необходима предварительная обработка речевого материала для учета неинформативных элементов речи, при создании языковых моделей, участвующих в процессе распознавания. Такая предварительная обработка позволяет более точно описать разговорную речь и исключить неинформативную часть дискурса.

2 Основные типы речевых сбоев спонтанной речи

Основные типы речевых сбоев можно разделить на два основных класса – *хезитации* или *колебания* и *самоисправления* [2]. Также в статье рассматриваются речевые сбои, связанные с *неправильным произношением* слов. Рассмотрим подробнее речевые сбои, влияющие на систему распознавания речи, на базе корпуса выступлений депутатов Верховной Рады Украины [5].

Распространёнными речевыми сбоями парламентариев являются: заполненные паузы одинарного и удвоенного звучания, слова, произнесенные неверно, фальстарты слов, повторы слов или словосочетаний, обрывы слов (недоговариваемые слова), коррекция со вставкой, он-лайн коррекция, повторы со вставкой и беспорядочные слова (имитация правильного произнесения слов).

Приведём краткое описание речевых сбоев, а также их влияние на процесс автоматического распознавания спонтанной украинской речи.

2.1 Заполненные паузы

Заполненные паузы – это так называемые вокализованные паузы. Они выполняют функцию заполнителя промежутков в речи. Такие промежутки речи возникают периодически, когда произношение текущей фразы завершено раньше, чем обдумано

(подготовлено) дальнейшее проистекание дискурса. Такого рода перерыв обычно называют *паузой колебания* или *паузой хезитации* [2]. Диктор использует заполненную паузу для обдумывания либо перепланирования следующей порции речи.

Таблица 1 – Примеры основных типов заполненных пауз

№	Тип речевого сбоя	Пример
1	Заполненные паузы одинарного звучания	[(**е**) привернуть (**е**) цю (**е**) сферу] [Що стосується (**е**) з ваших звинувачень]
2	Заполненные паузы двойного звучания	[наших громадян було в (**ее**) визволено] [але він (**ое**) вносить величезні зміни у]
3	Придыхания, откашливания	[складні проблеми для власне (**хе**)] [В мене питання. Чи ви (**кг**)]

Таким образом, употребление заполненных пауз позволяет избежать разрыва во фразе или диалоге с собеседником, образовавшаяся пауза заполняется различными звуками [3]. Примерами таких пауз могут быть:

– заполненные паузы, напоминающие фонемы («а», «е», «о») (п. 1);

– растягивание звуков «ее», «ме», «аа» («экание», «мэкание», «акание»), обладающих явной нефонологической долготой (п. 2);

– комбинации звуков «хе», «кг», «ги» (придыхания, откашливания) (п. 3).

Важнейшей функцией пауз является сегментация потока речи – маркировка границ между интонационными отрезками речи. Паузы на границах отрезков, образующих интонационное и семантическое единство, могут быть связаны как с членением потока речи, так и с колебаниями. Паузы внутри таких отрезков имеют, как правило, хезитационную природу. Кроме того, заполненные паузы являются частью самоисправлений диктора и могут находиться между забракованным элементом речи и его откорректированным фрагментом.

Следует отметить, что *заполненные паузы* являются наиболее распространённым видом речевых сбоев, присущих всем видам разговорной спонтанной речи.

2.2 Самоисправления

Внутренние проблемы развёртывания дискурса так же могут быть сведены ко второму классу речевых сбоев.

В определённый момент дискурса говорящий может решить, что некоторый фрагмент произнесенной им речи не соответствует изначальной формулировке мысли. Примером этому может быть несоответствующий или неточный подбор необходимых слов либо же некоторый фрагмент речи был произнесен преждевременно и не может в настоящем виде или в данном контексте быть адекватно введенным в дискурс. В этом случае диктор использует *самоисправления* [2].

Таблица 2 – Примеры основных типов самоисправлений

№	Тип речевого сбоя	Пример
4	Фальстарты слов	[процвітаючою, великою, (**поту**) **потужною** державою] [найбільш (**зне**) **знедолений** (**кат**) **категорій** населення.]
5	Коррекция со вставкой	[Шановні колеги! Я хочу (**попрос**) (**е**) (**попросити**)] [абсолютно не (**по**) **немає** (**потреби**) розглядати]
6	Он-лайн коррекция	[коли у видатках (**бур**) **бюджету**] [дуже (**турбо**) **стурбований**]

Примерами самоисправлений могут быть:

Фальстарты слов. Это один из видов коррекций, направленный на уточнение сказанного. В процессе произношения диктор произносит часть слова и в некоторый момент отказывается от этого фрагмента и полностью заменяет его (откорректированное слово содержит в себе последовательность букв забракованного фрагмента) (п. 4). Такого рода пробные шаги диктора, предшествующие произнесению правильного материала, часто называют *фальстартами* [2].

Коррекция со вставкой. Между забракованным фрагментом и его откорректированной частью могут возникать вставки, а именно: заполненные паузы (п. 5, пример 1, «е») или лексический материал (п. 5, пример 2, «немає»). Такая вставка может использоваться диктором для осмысления возможного варианта исправления предшествующего слова или фразы. После прерывания диктор заменяет забракованный фрагмент другим фрагментом (п. 5) или, как в случае фальстарта (п. 4), повторяет первоначально забракованный фрагмент, образуя правильное слово.

Он-лайн коррекция. В этом случае диктор немедленно реагирует на возникшую проблему в речи. Согласно «главному правилу прерывания» [2] диктор стремится остановить речевой поток в момент обнаружения проблемы, возможно, даже посредине слова. При коррекциях речевой отрезок до точки прерывания обладает незавершенными грамматическими, лексическими и просодическими характеристиками. Главное отличие *он-лайн коррекции* от *фальстарта* заключается в том, что забракованный фрагмент речи полностью отличается от последующего откорректированного варианта (п. 6, пример 2 «турбо» и откорректированный вариант – «стурбований»). В этом случае забракованный элемент речи не может быть охарактеризован как некоторая оборванная часть последующего самоисправления (п. 6).

Обрыв слов. В этом случае диктор не договаривает начатое слово без последующей коррекции (п. 7), либо делает обрыв слова и планирует его дальнейшее исправление (п. 4, п. 5, п. 6). В первом случае обрыв является результатом индивидуального произношения диктора (п. 7). Во втором случае диктор в точке прерывания речевого потока делает обрыв текущего фрагмента речи в любой его части. Из первого примера (п. 8) видно, что диктор, делая повтор словосочетания, оборвал последнее слово, а затем повторил и откорректировал забракованный фрагмент. Таким образом, обрывы слов могут быть составной частью других видов коррекций либо существовать как отдельный тип речевого сбоя, тем самым образовывая неинформативные фрагменты речи, которые не имеют своего представления в лексиконе системы автоматического распознавания.

Таблица 3 – Примеры основных типов самоисправлений

№	Тип речевого сбоя	Пример
7	Обрывы слов (недоговариваемые слова)	[наполягаєте ви на цій (**поправ...**).] [Шановні колеги! Нема потреби (**йо...**) обговорювати?]
8	Обрыв слов в повторе	[Але ветерани **самі** (**го...**) **самі говорять** зовсім про інше: вони проти.] [**платник податку** або не **платник** (**пода…**)]
9	Повторы слов или словосочетаний	[до залу**, який я** за **який я** доповідав] [**зауваження зауваження** стосувалися]
10	Повторы со вставкой	[до залу, (**який я**) **за** (**який я**) доповідав] [це (**без**) **то** (**без**) прийняття остаточного рішення]

Повторы слов. Представляют собой совмещение функции колебания и коррекции. Произнося фрагмент речи, диктор может остановиться, ощутив сомнение в том, что этот фрагмент соответствует изначальному ходу мыслей, затем отвергает эти сомнения и повторяет этот фрагмент заново. Повтор фрагмента свидетельствует о том, что диктор не подобрал более удачного слова или выражения, чем первоначальный фрагмент речи. Сомнения диктора в выборе подходящего фрагмента могут быть связаны с ранее произнесенным фрагментом, а именно: его формой, значением или местом расположения, а также с фрагментом, который находится в процессе обдумывания [2]. Так в п. 9 можно предположить, что диктор, делая повтор «зауваження зауваження», подбирал слово «стосувалися».

Повторы со вставкой. Представляет собой повтор фрагмента, разделенный заполненной паузой или фрагментом речи. Разделение повторов используется для выделения большего времени для переосмысления правильности употребления первого фрагмента и/или дальнейшего планирования дискурса (п. 10).

2.3 Неверно произнесенные слова

Слова, произнесенные неверно. Речевые сбои такого типа представляют собой слова, имеющие характерные черты редуцирования нескольких звуков. Такое редуцирование звуков, т.е. изменение звуков, состоящее в утрате полноты образования, приводит к появлению слов, смысл которых может быть обнаружен путём интеллектуального восстановления недостающих звуков с учетом контекста речи. Речевые сбои этого типа могут возникать в результате неполноты или неотчетливости артикуляции органов произношения (п.11).

Таблица 4 – Примеры основных типов самоисправлений

№	Тип речевого сбоя	Пример
11	Слова, произнесенные неверно	[(**епідоммічного**) благополуччя населення] [Я хотів би (**скати**) одну хвилинку]
12	Беспорядочные слова (имитация правильного произнесения слова)	[закон треба першому (**читні**) приймати] [двісті (**пітисізти**) тисяч осіб.]

Беспорядочные слова. Представляют собой «слова», смысл и значение которых может быть извлечен только из контекста произносимой речи. Причиной появления таких «слов» может послужить ускоренный темп речи и/или несоблюдение орфоэпических норм (п. 12).

2.4 Влияние речевых сбоев на процесс автоматического распознавания

Возникновение вокализованных пауз и фрагментов самоисправлений в речи диктора приводит к тому, что система распознавания не может найти соответствующее представление таких звуков в своём лексиконе. Эти новые звуки могут быть рассмотрены как слова, не известные системе распознавания («ее», «попрос», «турбо», «епідоммічного» и т.п.).

Наличие речевых сбоев подразумевает нарушение нормального потока слов. В результате последовательность слов, продуцируемая диктором, не в полной мере соответствует *языковой модели*, которая извлекается из текстового материала, не

содержащего речевых сбоев. Довольно часто декодер системы распознавания на месте речевого сбоя выбирает какое-то короткое функциональное слово, которое по своим акустическим свойствам подобно данному речевому сбою. В таком случае декодер системы распознавания будет генерировать неверную гипотезу распознавания слова, что в свою очередь может повлиять на возможность прогнозирования *языковой модели* в окрестности дисфлуенции. В результате можно ожидать, что в отдельных случаях один речевой сбой может быть причиной появления более чем одной ошибки на выходе системы распознавания [6].

Таким образом, речевые сбои оказывают негативное влияние на систему распознавания, настроенную на решение конкретной задачи, и должны быть либо учтены, либо удалены из исходного языкового материала.

3 Система автоматического распознавания слитной речи

В данной работе как базовая система используется инструментарий HTK [7] на основе скрытых Марковских моделей (СММ). Инструментарий HTK использовался для построения акустических и лингвистических моделей. Для распознавания речи был разработан программный комплекс [5], совместимый с акустическими и лингвистическими моделями HTK.

4 Акустическое и текстовое наполнение

4.1 Обучающая выборка

Обучение производилось на выступлениях депутатов Верховной Рады Украины, записанных через телевизионную сеть. Парламентская речь характеризуется некоторыми особенностями:

– Это спонтанная речь. Доклады парламентариев могут быть речью, прочитанной из заранее подготовленного текста доклада. Однако довольно часто происходит отступление от подготовленного доклада.

– Из-за ограничения во времени выступления многих дикторов произносятся в слишком быстром темпе.

– Часто речь эмоционально окрашена.

– Часто отсутствует ритмичность речи.

– Некоторые непоставленные голоса обладают тембральной немелодичностью: могут быть сиплыми, приглушенными или, наоборот, пронзительными, резкими.

– Основная часть записи состоит из непрерывных выступлений дикторов, изредка в них встречаются реплики ведущего заседания или других депутатов.

– Качество записи достаточно высокое, поскольку каждое депутатское место оснащено микрофоном.

Для обучения использовались записи длиной в 99 тыс. секунд, в которых встретилось 211 224 слова. Всего было записано 208 дикторов. Дикторов с длиной больше 300 с оказалось 87.

Обучение производилось на предварительно размеченной выборке. Для этого запись выступления автоматически разбивалась на фразы из нескольких слов, ограниченные паузами больше 400 мс. Каждой фразе оператором ставилась в соответствие метка в виде текста из стенограммы. Затем автоматически производилось преобразование текста в последовательность фонем в соответствии с контекстно-независимыми правилами. Выборка, размеченная таким образом, использовалась для построения акустической модели.

4.2 Анализируемая выборка

Процесс разметки, корректировки, анализа текста стенограммы в соответствии со звуковым сопровождением производился на стенограммах заседаний депутатов. Для анализа влияния основных неинформативных элементов речи на надёжность автоматического распознавания спонтанной речи была использована выборка из 11 файлов общей длительностью выступлений 35 390 секунд, в которых встретилось 73 775 слов и 203 докладчика. Для детального анализа были выбраны дикторы, которые произнесли в своём докладе не менее 500 слов, таких дикторов оказалось 46 человек, 8 дикторов из этой выборки выступали более одного раза. Время записи некоторых дикторов приведено в табл. 5.

Таблица 5 – Речевые сбои выборки из шести дикторов

Особенности диктора	GRY	ZVA	USH	TER	CIB	SAH	Сум. сбоев	Соотношение суммы сбоев к общ. числу сбоев 6 дикторов, %
Заполненная пауза «а»	4	12	4	47	22	25	114	40,86
Заполненная пауза «ее»	1	0	0	7	7	7	22	7,89
Фальстарты слов	0	1	1	2	4	0	8	2,87
Повторы	1	0	4	8	1	1	15	5,38
Обрывы	0	1	0	7	0	1	9	3,23
Коррекция со вставкой	0	1	3	4	4	1	13	4,66
Он-лайн коррекция	2	5	4	9	10	3	33	11,83
Повторы со вставкой	0	1	1	3	0	0	5	1,79
Слова, произнесенные неверно	20	3	14	12	6	1	56	20,07
Беспорядочные слова	0	0	1	1	1	1	4	1,43
Общ. кол. слов	660	513	663	1865	979	576	5256	-
Общ. кол. ошибок	28	24	32	100	55	40	279	-
Общ. кол. ошибок в докладе, %	4,24	4,68	4,83	5,36	5,62	6,94	5,31	-
Время выступления, с	354,88	252,06	305,79	874,05	448,30	287,03	2522,11	-
Темп, слов/с	1,86	2,04	2,17	2,13	2,18	2,01	-	-

4.3 Текстовый материал

Словарь был составлен из текстов стенограмм заседаний Верховной Рады Украины. С официального сайта Верховной Рады были загружены все стенограммы заседаний, начиная с 1991 года, что составило больше 100 МБ текста. Текст был модифицирован для того, чтобы убрать служебную информацию из стенограмм (например, аплодисменты), записать числа в текстовом виде, а также отделить русский текст от украинского.

Результирующий текст разделен на две части – первая содержит все тексты, кроме 2002 – 2003 годов, вторая содержит стенограммы 2002 – 2003 годов. Первая часть состоит из 14 629 111 слов, во второй содержится 409 244 слов.

5 Анализ характерных речевых сбоев группы дикторов

Для проведения анализа характерных речевых сбоев украинской речи дикторов и влияния этих сбоев на процесс автоматического распознавания речи была введена ручная разметка речевых данных. Подсчет расстройств речи конкретных дикторов и общей выборки дикторов реализован автоматически.

Табл. 5 представляет результаты анализа речевых сбоев выборки дикторов, которые в своём докладе сказали не менее 500 слов и совершили наибольшее количество речевых сбоев, где также указано общее число слов доклада, длительность доклада и темп произнесения каждого диктора.

Результаты изучения речи группы из 6 дикторов свидетельствуют, что наиболее повторяемыми нефонемными элементами от общего количества ошибок шести дикторов являются заполненные паузы «а» (40,86%), слова, произнесенные неверно (20,07%), он-лайн коррекция и коррекция со вставкой могут учитываться как один вид сбоя (16,49%) и заполненные паузы «ее» (7,89%). Для первых трёх дикторов, имеющих приблизительно одинаковое количество произнесенных слов, наблюдается увеличение количества ошибок в зависимости от повышения темпа речи.

Табл. 5 показывает, что число речевых сбоев, произнесенных каждым диктором в своей речи, варьируется от 4,24% до 6,94% от общего количества слов диктора. Для первых пяти дикторов частота появления речевых сбоев относительно числа слов, произнесенных диктором, возрастает.

В ходе корректировки и анализа текста стенограммы было обнаружено, что речевые сбои отдельно взятого диктора распределены неравномерно.

Разброс значений количества ошибок, совершаемых дикторами, существенно зависит от ораторских способностей отдельно взятого диктора.

6 Результаты экспериментов

В работе выполнена коррекция текста стенограммы в соответствии со звуковым сопровождением системы стенографирования выступлений депутатов Верховной Рады Украины. Проверка грамматики стенограммы осуществлялась с использованием электронного словаря «Словники України 3.0».

В процессе коррекции текста стенограммы в соответствии со звуковым сопровождением системы стенографирования были выявлены основные виды речевых сбоев, присущие спонтанной украинской речи. В текст стенограммы были внесены недостающие элементы спонтанной речи. Эти элементы были не учтены в лексиконе системы автоматического распознавания.

Экспертом была проведена разметка речевых данных, включающих элементы спонтанной речи депутатов. Разметка производилась для обнаружения условий и контекста появления речевых сбоев, оказывающих негативное влияние на систему автоматического распознавания речи. Главной особенностью введенной разметки текста стенограммы в соответствии со звуковым сопровождением является то, что такая разметка заключает в себе большую часть речевых сбоев для спонтанной украинской речи.

На базе собранного корпуса было проведено исследование нефонемных элементов спонтанной украинской речи. Проанализированы характерные ошибки укра-

инской речи, присущие конкретным дикторам (табл. 5). Результаты анализа речевых сбоев для 46 дикторов, которые сказали не менее 500 слов, контрольной выборки из 11 файлов записей заседаний приведены в табл. 6.

Таблица 6 – Речевые сбои выборки файлов

Файл	«а»	«ее»	Фальстарт	Повторы	Обрыв	Коррекции	Слова, произн. неверно	Беспор. слова	Общ. кол. сбоев	Общ. кол. слов	Сбои/общ. кол. слов, %	Время выступления	Средний темп, слов/с	Надёжн. распозн. реч. сбои, %	Надёжн. распозн. очищ.реч., %
10_16	27	15	3	3	1	11	27	5	92	5403	1,70	2512	2,19	76,07	79,61
09_26	29	24	4	10	8	16	40	1	132	2991	4,41	1369	2,19	82,94	84,26
10_08	42	17	2	8	9	41	108	12	239	7007	3,41	3354	2,11	78,55	81,72
10_18	43	11	5	8	9	28	73	7	184	7485	2,46	3394	2.19	69,93	72,9
10_22	69	29	6	9	7	37	63	5	225	11525	1,95	5978	1,89	73,71	73,95
10_23	55	13	9	7	1	16	14	3	118	4595	2,57	2205	2,03	74,62	74,95
10_24	99	13	10	36	35	49	92	6	340	9257	3,67	4114	2,17	76,28	77,16
10_25	63	16	6	11	7	15	20	5	143	5942	2,40	2794	2,17	63,83	64,02
11_12	16	6	7	3	18	18	29	9	107	8208	1,30	4280	1,94	70,82	71,19
11_19	35	5	2	2	1	6	11	1	63	4311	1,46	2004	2,18	74,99	75,44
11_20	45	21	6	8	4	30	38	8	160	7051	2,27	3386	2,09	73,89	74,17
Сумма сбоев по всей выборке	523	170	60	105	100	267	516	62	1803	73775	2,44	35390	-	-	-

Общее число речевых сбоев, отмеченное на выборке из 11 файлов, составляет 1803 единиц. Главное воздействие на результат распознавания оказывают заполненные паузы одинарного звучания, их доля составляет 28,95% (от общего числа речевых сбоев), слова, произнесенные неверно, составляют 28,62%, а также все виды коррекций, включая фальстарты слов, составляют 18,14%. Заполненные паузы с растягиванием звуков («ее», «ме», «аа» и т.п.) составляют 9,43%. Остальные типы сбоев не превышают 5,82%.

Тем не менее, в своей совокупности все виды аннотируемых речевых сбоев оказывают негативное влияние на систему распознавания речи. Исключением могут стать только четко артикулируемые повторы (5,82%). В ходе анализа результатов распознавания и текста стенограммы в соответствии со звуковым сопровождением не было обнаружено существенного влияния артикулируемых повторов на надёжность автоматического распознавания. Поэтому для подряд стоящих артикулируемых повторов одного слова разметка не вводилась. В случае возникновения недоартикулированных повторов неинформативный фрагмент повтора относился к классу обрывов.

Количество речевых сбоев, произнесенных 46 дикторами, составляет 2,44% относительно всех слов, произнесенных 46 дикторами.

Несмотря на относительно незначительный вклад речевых сбоев в спонтанную украинскую речь, практика показала, что содержание речевых сбоев речи следует исследовать для каждого диктора в отдельности. Количество сбоев каждого анализируемого файла отличается в зависимости от того, какие дикторы попали в выборку. При этом число произнесенных речевых сбоев относительно всех слов отдельно взятого диктора, попавшего в контрольную выборку, изменяется от 0,39 % до 6,94%.

В работе проведён учет и автоматическое удаление сегментов речи, которые содержат любой из вышеуказанных типов сбоев, исключением были только четко артикулируемые повторы, которые не создавали «мусорных» неинформативных элементов. Результат проведения эксперимента по распознаванию анализируемой выборки украинской спонтанной речи на материалах, очищенных от речевых сбоев, дал повышение надёжности распознавания в среднем 1,25% (табл. 6).

Надежность распознавания сильно отличается в зависимости от того, какие дикторы попали в выборку. Например, в файл 09_26 попал доклад только одного диктора, который произнёс 4,41% речевых сбоев относительно всех слов своей речи. В выборку 10_16 попало 6 докладчиков, у которых число произнесенных речевых сбоев изменялось от 0,75% до 3,72%.

Наилучший результат проведения эксперимента по распознаванию украинской спонтанной речи на материалах, очищенных от речевых сбоев, был получен для анализируемых выборок 10_16 и 10_08 и дал повышение надёжности распознавания 3,5% и 3,2% соответственно.

Таблица 7 – Результаты распознавания контрольной выборки

Файл	Длина, с	Надёжность распознавания со сбоями, %	Надёжность распознавания, очищенная речь, %	Изменение, %
09_10а	4925	66,22	71,46	5,24
09_27	1521	74,40	81,19	6,79
10_15	8678	70,56	77,30	6,74
10_15а	5022	71,29	78,85	7,56
10_17	6088	79,09	84,37	5,28
10_17а	3282	76,65	82,60	5,95
11_21	7433	75,79	83,41	7,62

Так же проведены эксперименты по распознаванию слитной спонтанной речи на контрольной выборке, в которую вошли 7 файлов записей заседаний, на которых не производилось обучение системы распознавания речи (табл. 7).

Была произведена дополнительная коррекция и проверка грамматики стенограммы контрольной выборки в соответствии со звуковым сопровождением системы стенографирования [5]. Проверка грамматики осуществлялась с использованием электронного словаря «Словники України 3.0». Экспертом была проведена разметка речевых данных контрольной выборки, включающих элементы спонтанной речи депутатов. Детальный анализ речевых сбоев контрольной выборки не проводился. Последняя колонка показывает изменение надёжности распознавания для очищенной от речевых сбоев речи. Результат проведения распознавания на контрольной выборке дал повышение надёжности распознавания в среднем на 6,5%.

Заключение

Полученные данные и результаты могут быть в дальнейшем использованы для разработки подходов учета структуры и моделирования элементов, присущих спонтанной украинской речи. Автоматизирование процесса выявления и разметки речевых сбоев позволит уменьшить долю ручного аннотирования и охватить больший круг описываемых явлений.

Литература

1. Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop. 10 – 12 September 2005, Aix-en-Provence, France.
2. Подлесская В.И. Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи / В.И. Подлесская, А.А. Кибрик // Научно-техническая информация. – Серия 2. – 2007. – № 2. – С. 2-23.
3. Леонтьева Ал.Б. Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной речи / Ал.Б. Леонтьева, И.С. Кипяткова // Труды Первого междисциплинарного семинара «Анализ разговорной русской речи» AR3 – 2007. – СПИИРАН. – Санкт-Петербург, 2007. – С. 77-85.
4. Zechner Klaus. Automatic Generation of Concise Summaries of Spoken Dialogues in Unrestricted Domains / Klaus Zechner // In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. – New Orleans, Louisiana, USA. – 2001, September. – P. 199-207.
5. Пилипенко В.В. Автоматизированный стенограф украинской речи / В.В. Пилипенко, В.В. Робейко // Штучний інтелект. – 2008. – № 4. – С. 768-775.
6. Stouten F. Coping with disfluencies in spontaneous speech recognition: Acoustic detection and linguistic context manipulation / F. Stouten, J. Duchateau, J.P. Martens [и др.] // Speech Communication. – 2006. – Vol. 48. – P. 1590-1606.
7. The HTK Book / Young S., Evermann G., Kershaw D. [и др.] – Cambridge University Engineering Department, 2002.

В.В. Пилипенко, О.М. Ладошко

Анотація та врахування мовних збоїв у задачі автоматичного розпізнавання сонтанної української мови

У статті розглядається проблема впливу мовних збоїв на процес автоматичного розпізнавання спонтанної української мови на прикладі автоматизованого стенографа для отримання тексту стенограми із звукового файла. Записана фонограма обробляється системою розпізнавання злитого мовлення багатьох дикторів з великих словників (більше 10 тис. слів). Розглядаються і систематизуються основні типи мовних збоїв спонтанної української мови. На основі врахування мовних збоїв та корекції стенограми робиться очищення даних, що дозволяє покращувати показники надійності розпізнавання мови.

V.V. Pylypenko, O.N. Ladoshko

Annotation and Accounting Disfluencies in the Problem of Automatic Recognition of Ukrainian Speech

The paper is about the problem of disfluencies effect on the processing of automatic recognition of spontaneous Ukrainian speech. It is considered by example of a computerized stenographer. It makes the text from sound records based on the speech recognition system aided by human. Large vocabulary (more than 10K words) continuous speech recognition system for a number of speakers is used to process recorded files. The paper investigates and classifies the basic types of disfluencies in spontaneous Ukrainian speech. On the basis of disfluencies annotation and transcript correction the cleaning of the speech data was provided. It was made to improve the recognition rate in automatic speech recognition possible.

Статья поступила в редакцию 19.07.2010.